

Combining Labeled and Unlabeled Data with Co-Training

(1998) (Make Corrections) (180 citations)

Avrim Blum, Tom Mitchell

COLT: Proceedings of the Workshop on Computational Learning Theory, Morgan Kaufmann Publishers

^c_u [Bookmark in CiteULike](#)



[Home/Search](#) [Bookmark](#) [Context](#) [Related](#)

View or download:

cmu.edu/project/theo...colt98_final.ps
cmu.edu/pub/files/...lum_a_1998_1.ps.gz
cmu.edu/project/theo...colt98_final.ps

Cached: [PS.gz](#) [PS](#) [PDF](#)

[Image](#) [Update](#) [Help](#)

From: cmu.edu/~webkb/ (more)

Homepages: [A.Blum](#) [T.Mitchell](#)

Links: [ACM](#)

([Enter summary](#))

Rate this article: 1 2 3 4 5 (best)

[Comment on this article](#)

Abstract: We consider the problem of using a large unlabeled sample to boost performance of a learning algorithm when only a small set of labeled examples is available. In particular, we consider a problem setting motivated by the task of learning to classify web pages, in which the description of each example can be partitioned into two distinct views. For example, the description of a web page can be partitioned into the words occurring on that page, and the words occurring in hyperlinks that point to... ([Update](#))

Cited by: [More](#)

A Benchmark Dataset for Audio Classification and.. - Homburg, Mierswa.. (2005) ([Correct](#))

Adaptive View Validation: A First Step Towards Automatic.. - Ion Muslea Muslea ([Correct](#))

Journal of Machine Learning Research 3 (2003) 1183-1208.. - Ron Bekkerman Ronb ([Correct](#))

Active bibliography (related documents): [More](#) [All](#)

0.9: Combining Labeled and Unlabeled Data with Co-Training - Blum, Mitchell (1998) ([Correct](#))

0.3: PAC Learning from Positive Statistical Queries - Denis (1998) ([Correct](#))

0.1: Multi-Document Summarization and Visualization in the Informedia.. - Wactlar (2001) ([Correct](#))

Similar documents based on text: [More](#) [All](#)

0.6: Learning from Labeled and Unlabeled Data using Graph Mincuts - Blum, Chawla (2001) ([Correct](#))

0.4: Learning From A Mixture Of Labeled And Unlabeled Examples.. - Joel Ratsaby And (1995) ([Correct](#))

0.2: A PAC-style Model for Learning from Labeled and Unlabeled Data - Balcan, Blum (2004) ([Correct](#))

Related documents from co-citation: [More](#) [All](#)

30: Text classification from labeled and unlabeled documents using EM - Nigam, McCallum et al. - 1999

23: Maximum Likelihood from Incomplete Data via the EM Algorithm (context) - Dempster, Laird et al. - 1977

22: Text categorization with Support Vector Machines: Learning with many relevant fe.. - Joachims - 1998

BibTeX entry: ([Update](#))

Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training.

<http://citeseer.ist.psu.edu/blum98combining.html> [More](#)

```
@inproceedings{ blum98combining,
  author = "Avrim Blum and Tom Mitchell",
  title = "Combining Labeled and Unlabeled Data with Co-training",
  booktitle = "{COLT}: Proceedings of the Workshop on Computational Learning Theor",
  pages = "92-100",
  year = "1998",
  url = "citeseer.ist.psu.edu/blum98combining.html" }
```

Citations (may not include all citations):

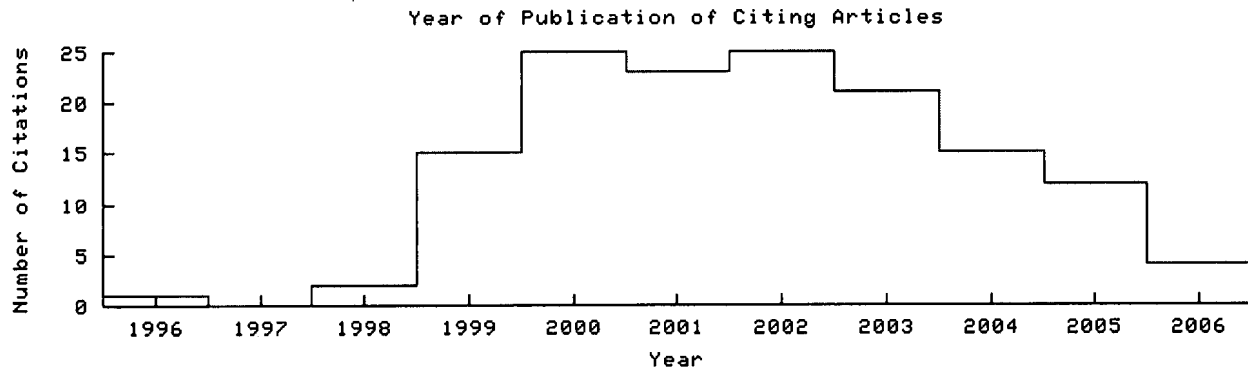
2528 Maximum likelihood from incomplete data via the EM algorithm (context) - Dempster, Laird et al. - 1977

2133 Pattern Classification and Scene Analysis (context) - Duda, Hart - 1973

149 Learning to extract symbolic knowledge from the world wide w.. - Craven, Freitag et al. - 1997 [ACM](#) [DBLP](#)

110 Unsupervised word sense disambiguation rivaling supervised m.. - Yarowsky - 1995 [ACM](#) [DBLP](#)

- 107 Efficient noise-tolerant learning from statistical queries - Kearns - 1993 [ACM](#) [DBLP](#)
- 97 A comparison of two learning algorithms for text categorizat.. - Lewis, Ringuette - 1994
- 76 Supervised learning from incomplete data via an EM approach - Ghahramani, Jordan - 1994 [DBLP](#)
- 38 Informedia: News-on-demand - multimedia information acquisit.. - Hauptmann, Witbrock - 1997
- 19 A computational model of teaching - Jackson, Tomkins - 1992
- 19 Learning from a mixture of labeled and unlabeled examples wi.. - Ratsaby, Venkatesh - 1995 [ACM](#) [DBLP](#)
- 17 and network design problems (context) - Karger, in et al. - 1994
- 17 and network design problems (context) - Karger, in et al. - 1997
- 14 Journal of Computer and System Sciences (context) - Goldman, Kearns et al. - 1995
- 7 Improving acoustic models by watching television - Witbrock, Hauptmann - 1998
- 3 PAC learning with constantpartition classification noise and.. (context) - Decatur - 1997



The graph only includes citing articles where the year of publication is known.

Documents on the same site (<http://www.cs.cmu.edu/~webkb/>): [More](#)

First-Order Learning for Web Mining - Craven (1998) ([Correct](#))

A Case Study in Using Linguistic Phrases for Text.. - Fürnkranz, Mitchell.. (1998) ([Correct](#))

Using Reinforcement Learning to Spider the Web Efficiently - Rennie, McCallum (1999) ([Correct](#))

[Online articles have much greater impact](#) [More about CiteSeer.IST](#) [Add search form to your site](#) [Submit documents](#) [Feedback](#)

CiteSeer.IST - Copyright [Penn State](#) and [NEC](#)

CiteSeer.IST Home **Correcting:** Text Classification by Bootstrapping with Keywords, EM and.. - McCallum, Nigam (1999) (Correct)

Correct or update the information below. **All corrections are logged and reviewed. Malicious changes and other hacking attempts are specifically prohibited.** Correct or update title and author information.

Abstract (or
first
paragraph):

When applying text classification to complex tasks, it is tedious and expensive to hand-label the large amounts of training data necessary for good performance. This paper presents an alternative approach to text classification that requires no labeled documents; instead, it uses a small set of keywords per class, a class hierarchy and a large quantity of easily obtained

One line
summary:

(Summarize the contributions of this paper. Why is this paper important/useful)
Papers with one line summaries are highlighted on the CiteSeer.IST homepage.
The title is always shown with the summary, so there is no need to repeat the title

BibTeX
entry:

```
@misc{ mccallum99text,
  author = "A. McCallum and K. Nigam",
  title = "Text Classification by Bootstrapping with Keywords",
  text = "A. McCallum, K. Nigam, Text Classification by Bootstrapping
with Keywords,
      EM and Shrinkage, ACL Workshop for Unsupervised Learning in
Natural Language
      Processing, 1999",
```

Author
homepages:

(Syntax: First Initial Surname = URL, e.g. J Smith = <http://jsmith.org/> K der Smith = <http://kdersmith.com/>. May be used to override or add author homepage links. If needed, the author field must also be updated.)

Submit Correction